11-1959

Smoothing in Discrete Servo-Stochastic Processes

Harlan D. Mills

SMOOTHING IN DISCRETE SERVO-STOCHASTIC PROCESSES

Harlan D. Mills

MATHEMATICA

92-A Nassau Street *Princeton, New Jersey*

Subsidiary of MARKET RESEARCH CORPORATION OF AMERICA

SMOOTHING IN DISCRETE SERVO-STOCHASTIC PROCESSES

Harlan D. Mills

November 1959

101-02

MATHEMATICA

Princeton, N. J.

Subsidiary of MARKET RESEARCH CORPORATION OF AMERICA

ABSTRACT

SMOOTHING IN DISCRETE SERVO-STOCHASTIC PROCESSES

An "uncertainty principle" is deduced in discrete servo-stochastic processes satisfying a conservation relation

$$e_{t+1} = e_t + c_{t-T} + n_t$$

for errors e_t with independent identically distributed noise n_t and servo corrections c_{t-T} lagged T periods. Ratios of "error to noise", $k_e = \sigma_e / \sigma_n$, and "correction to noise", $k_c = \sigma_c / \sigma_n$, are shown to satisfy a <u>smoothing capacity</u> relation

$$k_e^2 \geq \frac{1}{4}(k_c + \frac{1}{k_c})^2 + T,$$

for any servo with bounded mean errors (and in particular, $k_e k_c > \frac{1}{2}$). Thus not both k_e and k_c can be decreased indefinitely by improving servo design.

A family of servos,

$$c_{t+1} = - \alpha e_t + \beta, \; 0 < \alpha \leq 1, \quad \beta \text{ arbitrary},$$

is shown to be a complete optimal class, providing maximal smoothing (achieving the equality above). That so simple a class of servos is optimal appears rather fortuitous in view of its competition with all possible servos with bounded mean errors.

SMOOTHING IN DISCRETE SERVO-STOCHASTIC PROCESSES

Harlan D. Mills

General Theory

We define a discrete servo-stochastic process to be a sequence of random variables [2, 4, 5]

$$P = (e_1, c_1, n_1, e_2, c_2, n_2, \ldots,) \quad ,$$

where we interpret

e_t as an "error" in an operation at the beginning of a time period t,
c_t as a "correction" made in the operation at the beginning of period t,
n_t as a "noise" entering the operation during period t,

such that P, in conjunction with a squence of numbers,

$$I = (\ldots, e_{-1}, c_{-1}, n_{-1}, e_0, c_0, n_0) \quad ,$$

called the "initial conditions", is characterized by the following properties, for $t = 0, 1, 2, \ldots,$

(1) $\quad e_{t+1} = e_t + c_t + n_t$

(2) $\quad c_{t+1} = S(\ldots, n_{t-1}, e_t, c_t, n_t)$

(3) $\quad \text{Prob}\ \{n_t \leq x\} = N(x)$

where S is an arbitrary function and N is an arbitrary distribution. Equation (1) represents a conservation of error in the operation (holding with probability 1), (2) specifies a "servo" or "decision policy" S, (holding with probability 1), and (3) describes noise - taken to be identically and independently distributed period by period.

It is clear that conditions (1), (2), and (3), with the initial conditions, I, uniquely determine the process P. for (1), (2), (3) can be applied recursively to determine the distribution of any partial sequence

$$P_k = (e_1, c_1, n_1, e_2, c_2, n_2, \ldots, e_k, c_k, n_k) \quad .$$

Illustrations which motivate the formulation and study of such processes include multi-echelon inventory supply systems, multi-stage manufacturing systems, space tracking systems, guidance and prediction systems, vehicle convoy systems, and computational programs. Arrow, Karlin and Scarf [1], Brown [3], Pinkham [6], Vassian [7], and Wiener [8], have been concerned with closely related problems.

Means and variances in P will be written as

$$\bar{e}_t = E(e_t), \; \bar{c}_t = E(c_t), \bar{n}_t = E(n_t)$$

$$\sigma_e^2(t) = E[(e_t - \bar{e}_t)^2], \text{ etc.}$$

P is said to be a stable process if all first and second order moments approach definite limits in time; we define

$$\bar{e} = \lim_{t \to \infty} \bar{e}_t \, , \quad \bar{c} = \lim_{t \to \infty} \bar{c}_t \, , \quad \bar{n} = \bar{n}_t$$

$$\sigma_e^2 = \lim_{t \to \infty} \sigma_e^2(t), \text{ etc.}$$

We shall be concerned only with stable processes - excluding, for example, servos which allow the error to become unbounded. In stable processes we shall be interested in two ratios - "error to noise" and "correction to noise" - defined as

$$(4) \qquad k_e = \sigma_e / \sigma_n , \qquad k_c = \sigma_c / \sigma_n \quad .$$

We shall assume, hereafter, that $\sigma_n^2 > 0$. In many circumstances, it is desired to design a servo to keep both of these ratios small - the first gives a measure of "how well the servo is doing", and the second of "how hard the servo is working."

First, we shall develop an "uncertainty principle" for servo-stochastic processes, namely that for any servo,

$$(5) \qquad k_e \geq \frac{1}{2}(k_c + \frac{1}{k_c}) ,$$

or, in another form,

$$(6) \qquad k_e k_c \geq \frac{1}{2}(1 + k_c^2) \geq \frac{1}{2} ,$$

which demonstrates that both ratios cannot be decreased indefinitely by better servo designs. We call this relation the "smoothing capacity" of servos in discrete servo-stochastic processes.

Next, we formulate a complete optimal class of servos which achieves the equation form of the inequality of (5). These optimal servos turn out to be linear, namely, of the form

$$(7) \qquad c_t = \gamma - \alpha(e_t - \varepsilon)$$

where $0 < \alpha \leq 1$, and γ, ε are arbitrary. That so simple a class of policies is complete and optimal seems quite fortuitous considering the fact that they are in competition with all possible servos, nonlinear, discontinuous, or what have you.

Finally, we formulate and characterize the effect of information time delays and/or servo-mechanical response times on these ratios and on the smoothing capacity. With a composite time delay of T periods, the "uncertainty principle" is amended to

$$(8) \qquad k_e^2 \geq \frac{1}{4}(k_c + \frac{1}{k_c})^2 + T$$

and the class of policies given by (7) remains the complete optimal class.

Smoothing Theorems

Two theorems follow which develop results on smoothing in servo-stochastic processes. Theorem 1 (Smoothing Capacity) determines the boundary of (5) and thereby the smoothing capacity of servos (with no information or mechanical delay). Theorem 2 (Optimal Servo Class) establishes the completeness and optimality of the class of servos given by (7). Other equivalent forms of this servo class are also presented.

THEOREM 1. (Smoothing Capacity). In a stable discrete servo-stochastic process, for any servo whatsoever,

$$(5) \quad k_e \geq \frac{1}{2}\left(k_c + \frac{1}{k_c}\right)$$

Proof. By (1)

$$e_{t+1} = e_t + c_t + n_t$$

and

$$\bar{e}_{t+1} = \bar{e}_t + \bar{c}_t + \bar{n}_t \quad ;$$

or,

$$(e_{t+1} - \bar{e}_{t+1}) = (e_t - \bar{e}_t) + (c_t - \bar{c}_t) + (n_t - \bar{n}_t)$$

Square both sides of this latter equation, take expectations, and then limits to obtain

$$\sigma_e^2 = \sigma_e^2 + \sigma_c^2 + \sigma_n^2 + 2\gamma_{ec}\sigma_e\sigma_c$$

where γ_{ec} is the correlation between e_t and c_t as $t \to \infty$, $-1 \leq \gamma_{ec} \leq 1$, and the other two cross product terms of the right side are zero since n_t is independent of e_t and c_t. Using the definitions of k_e and k_c, this last equation can be rewritten as

$$[k_c^2 + 1 + 2\gamma_{ec} k_e k_c] \, \sigma_n^2 = 0$$

which can be rearranged, since $\sigma_n^2 \neq 0$, as

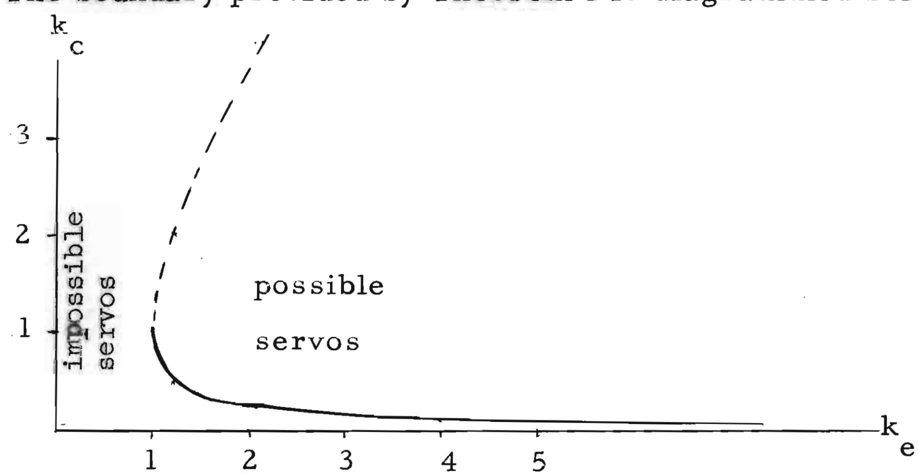
$$(9) \qquad -\gamma_{ec} k_e = \frac{1}{2}\left(k_c + \frac{1}{k_c}\right)$$

Now $-\gamma_{ec} > 0$ necessarily since $k_e > 0$ and the right side of this equation is positive for any $k_c > 0$. Furthermore, $-\gamma_{ec} \leq 1$, so

$$k_e \geq -\gamma_{ec} k_e = \frac{1}{2}\left(k_c + \frac{1}{k_c}\right)$$

as was to be proved.

The boundary provided by Theorem 1 is diagrammed below



We have drawn the boundary of interest as a solid curve, the remainder as a dotted curve. The solid curve represents the barrier to the joint minimization of the ratios k_e and k_c. Notice that everywhere on this barrier, $k_c = \alpha k_e$ for some α such that $0 < \alpha \leq 1$. Theorem 2 will establish an optimal class of servos which generates this entire barrier.

It is an interesting side point to note that given a certain error to noise ratio k_e, say, not only is there a lower bound to possible correction to noise ratios k_c, but there is an upper bound as well - and these lower and upper bounds are reciprocals. Thus, if servos can only do so well in this respect, they can also only do so badly, and still provide a stable process.

Curiously enough, it will be apparent, from the proof of the next theorem that this "worst possible" servo has the same form of the optimal class, but for a different range of the parameter α - namely, the class given by (7), except that $1 \leq \alpha < 2$, sweeps out the dotted curve. Thus all servos which provide a given error to noise ratio have correction to noise ratios which lie between a certain number and its reciprocal.

THEOREM 2. (Optimal Servo Class). Given α such that $0 < \alpha \leq 1$, the servo

$$(7) \quad c_t = \gamma - \alpha(e_t - \varepsilon)$$

where γ and ε are arbitrary constants, determines a stable discrete servo-stochastic process such that

$$(5') \quad k_e = \frac{1}{2}(k_c + \frac{1}{k_c}) \text{ and } k_c = \alpha k_e$$

Proof. We establish, first, that (7) determines a stable process, and then that it has the required properties. To accomplish this we exhibit a solution to the pair of difference equations

$$(1) \quad e_{t+1} = e_t + c_t + n_t$$

$$t = 0, 1, 2, \ldots,$$

$$(7) \quad c_{t+1} = \gamma - \alpha(e_{t+1} - \varepsilon)$$

with e_0, c_0, n_0 being given numbers. Substitute (7) in (1) to obtain

$$e_{t+1} = (1-\alpha)e_t + \gamma + \alpha\varepsilon + n_t$$

with an easily verified solution

$$(10) \quad e_t = \frac{\gamma}{\alpha} + \varepsilon + \sum_{j=0}^{t-1} (1-\alpha)^j n_{t-1-j} + (1-\alpha)^t e_0, \quad t = 1, 2, \ldots,$$

and on resubstituting (10) into (7),

$$(11) \quad c_t = -a \sum_{j=0}^{t-1} (1-\alpha)^j n_{t-1-j} - \alpha(1-\alpha)^t e_0 \ .$$

Now $|1-\alpha| < 1$ when $0 < \alpha \leq 1$ and (10) and (11) certainly exhibit a stable process, since the noise, n_j, is identically and independently distributed.

We notice further, from (7), that

$$\bar{c}_{t+1} = \gamma - \alpha(\bar{e}_{t+1} - t),$$

so that

$$(c_t - \bar{c}_t) = -\alpha(e_t - \bar{e}_t) \quad . \tag{12}$$

Squaring both sides of (12), and taking expectations and limits, gives

$$\sigma_c^2 = \alpha^2 \sigma_e^2 , \text{ or } \sigma_c = \alpha\sigma_e$$

and, hence,

$$k_c = \alpha k_e \quad .$$

On the other hand, multiplying both sides of (12) by $(e_t - \bar{e}_t)$, and taking expectations and limits, gives

$$\gamma_{ec} \sigma_e \sigma_c = -\alpha\sigma_e^2 \quad \text{or} \quad \gamma_{ec} \sigma_e = -\alpha\sigma_e$$

and hence, since $\sigma_e > 0$,

$$\gamma_{ec} = -1.$$

But referring to equation (9) in the proof of Theorem 1, we find that if $\gamma_{ec} = -1$ then

$$k_e = \frac{1}{2}(k_c + \frac{1}{k_c}) \quad ,$$

as was to be proved. Thus we have shown (7) determines a stable process and leads to both properties stated in (5'). This completes the proof of the Theorem.

Incidentally the explicit forms of e_t and c_t in (10) and (11) allow us to compute $\bar{e}$, $\bar{c}$, σ_e^2, σ_c^2 independently, as follows

$$\bar{e} = \varepsilon + \frac{1}{\alpha}(\gamma + \bar{n}), \quad \bar{c} = -\bar{n}$$

$$\sigma^2 = \sum_{j=0}^{\infty} (1-\alpha)^{2j} \sigma_n^2 = \frac{1}{1-(1-\alpha)^2} \sigma_n^2$$

$$\sigma_c^2 = \alpha^2 \sum_{j=0}^{\infty} (1-\alpha)^{2j} \sigma_n^2 = \frac{\alpha^2}{1-(1-\alpha)^2} \sigma_n^2$$

whence

$$(13) \qquad k_e^2 = \frac{1}{1-(1-\alpha)^2}, \quad k_c^2 = \frac{\alpha^2}{1-(1-\alpha)^2}$$

Note, in particular, that k_e and k_c are independent of the arbitrary choice of γ and ε in (7). In fact, though there are, formally, three parameters, α, γ, ε, in (7), only two parameters, say α and $\beta = \frac{\gamma}{\alpha} + \varepsilon$ determine the process, and only α affects the error to noise and correction to noise ratios.

The optimal servo class given by

$$(7) \qquad c_t = \gamma - \alpha(e_t - \varepsilon)$$

has, as we have seen, the alternative representation of

$$(11) \qquad c_t = -\alpha \sum_{j=0}^{t-1} (1-\alpha)^j n_{t-1-j} - \alpha(1-\alpha)^t e_0 ,$$

and (eliminating e_{t+1} and e_t from (1) by means of (7)), still another form is

$$(13) \qquad c_t = (1-\alpha)c_t - \alpha n_t .$$

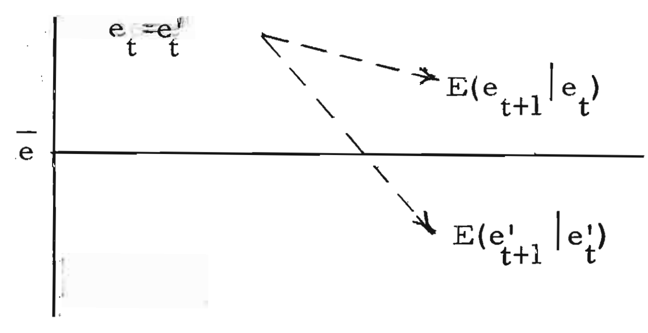
Additional insight into the servo of (7) comes from examining

$$\begin{aligned} E(e_{t+1} | e_t) &= E(e_t + c_t + n_t | e_t) \\ &= E(e_t + \gamma - \alpha(e_t - \varepsilon) + n_t | e_t) \\ &= (1-\alpha) e_t + \gamma + \alpha\varepsilon + \bar{n} \\ &= (1-\alpha)e_t + \alpha \bar{e} . \end{aligned}$$

Thus, the servo always seeks, in expectation, to move a fraction α of the way from e_t to $\bar{e}$ - another behaviorial characteristic of the servo which is independent of the parameters γ and ε. On the other hand, in the "worst possible" servo, obtained by letting $1 \leq \alpha < 2$ in (7), identify $\alpha' = 2 - \alpha$, so $0 < \alpha' \leq 1$, the same range as α. In the resulting process, we compute

$$
\begin{aligned}
E(e'_{t+1} \mid e'_t) &= E(e'_t + c'_t + n'_t \mid e'_t) \\
&= E(e'_t + \gamma - (2-\alpha')(e' - \varepsilon) + n'_t \mid e'_t \\
&= -(1-\alpha')e'_t + \gamma + (2-\alpha')\varepsilon + \bar{n} \\
&= -(1-\alpha')e' + (2-\alpha')\overline{e'} \\
&= (1-\alpha)\,(-e'_t) + \alpha\,\overline{e'}
\end{aligned}
$$

Thus, this servo always seeks, in expection, to move a fraction α of the way from $-e'$ to $\overline{e'}$. We also note that when $\alpha' = 2-\alpha, (1-\alpha)^2 = (1-\alpha')^2$, so, referring to (13), $k_e = k_{e'}$ for corresponding α and α'. If the two processes have a common mean error, $\bar{e} = \overline{e'}$ (which occurs when $\bar{c} = -\bar{n}$) these relations can be summarized as follows. Given any possible error to noise ratio, $k_e \geq 1$, there exists α and $\alpha' = 2 - \alpha$ such that the respective processes determined by (7) have the given error to noise ratio. The expectations of the correction processes are diagrammed below.



Thus the variability of the errors about the mean is identical in the two processes, but the mode of correction is quite different. The corrections diagrammed represent the behavior at $e_t = e'_t$ of the "best" and "worst" possible servos with a given error to noise ratio.

Time Dependent Smoothing Theorems

The following theorems extend the results of Theorems 1 and 2 to situations with time delays in the servo processes. Theorem 3 (Information Delay) formulates the effect of non zero information time delays and/or servo-mechanical response times on servos. Theorem 4 summarizes and organizes the first three theorems into a comprehensive statement to obtain the time dependent smoothing capacity of (8) and verification of the completeness and optimality of the class of servos given by (7) under any such condition of information delay.

Up to this point, we have assumed that corrections in the process are made instantaneously. Suppose, instead that servo-mechanical considerations require T periods for a desired correction to be realized. Then (1) should read

$$(1^T) \qquad e_{t+1} = e_t + c_{t-T} + n_t$$

On the other hand, suppose every item of information in the process is delayed by T periods. Then (2) should read

$$(2^T) \qquad c_{t+1} = S(\ldots, n_{t-T-1}, e_{t-T}, c_{t-T}, n_{t-T})$$

It is easy to verify, formally, that the two sets of properties (1^T), (2), (3) and (1), (2^T), (3) determine identical stochastic processes. Thus delays in servo-mechanical responses and delays in information, or any combination of them, are equivalent. For convenience, then we can simply phrase all such delays as information delays. The effects of such delays are characterized in the following theorem.

THEOREM 3. (Information Delay) Let S^T be identical to servo S except that information is delayed T periods in its use by S^T, i.e.,

$$S^T(\ldots, n_{t-1}, e_t, c_t, n_t) = S(\ldots, n_{t-T-1}, e_{t-T}, c_{t-T}, n_{t-T})$$

then if S determines a stable discrete servo-stochastic process, S^T also determines a stable process with the properties

$$(14) \quad [k_e(S^T)]^2 = [k_e(S)]^2 + T, \; k_c(S^T) = k_e(S) .$$

Proof. Let $P = \{e_t, c_t, n_t\}$ be the stable discrete servo-stochastic process determined by S. Then

$$(1) \qquad e_{t-T+1} = e_{t-T} + c_{t-T} + n_{t-T}$$

Add, now, the T terms, $n_{t-T+1} + \ldots + n_t$, to both sides of this equation, and regroup, as

$$(e_{t-T+1} + n_{t-T+1} + \ldots + n_t) = (e_{t-T} + n_{t-T} + \ldots + n_{t-1}) + c_{t-T} + n_t$$

Then $P^T = \{e_t^T, c_t^T, n_t\}$, where

$$e_t^T = e_{t-T} + n_{t-T} + \ldots + n_{t-1}, \; c_t^T = c_{t-T}$$

satisfies (1) and

$$(2^T) \qquad c_{t+1}^T = S^T(\ldots, n_{t-1}, e_t^T, c_t^T, n_t)$$; hence P^T is the process determined by S^T. Now, it is clear that P^T is stable if P is. In addition the values of σ_e^2 and σ_n^2 may be computed directly from the form of P^T, and are as given. This completes the proof of the Theorem.

Finally, we note that the results of the three preceding Theorems can be organized into a single comprehensive statement about discrete servo-stochastic processes as follows.

THEOREM 4. In a stable discrete servo-stochastic process with information time delay T, for any servo whatsoever,

$$(8) \quad k_e^2 \geq \frac{1}{4}(k_c + \frac{1}{k_c})^2 + T \quad .$$

The servo

$$(7) \quad c_t = \gamma - \alpha(e_t - \varepsilon), \quad 0 < \alpha \leq 1,$$

where γ and ε are arbitrary constants, determines a stable process such that

$$(8') \quad k_e^2 = \frac{1}{4}(k_c + \frac{1}{k_c})^2 + T;$$

Furthermore, the class of servos given by (7) generates the entire minimal boundary of (8) as α varies from 0 to 1.

Bibliography

[1] Arrow, K. J., S. Karlin and H. Scarf. Studies in the Mathematical Theory of Inventory and Production, Stanford, 1958.

[2] Bartlett, M. S. An Introduction to Stochastic Processes, Cambridge, 1956.

[3] Brown, W. M. Analysis of Discrete Linear Systems. J. SIAM. Vol. 5, 1957.

[4] Doob, J. L. Stochastic Processes, Wiley, 1953.

[5] Feller, W. An Introduction to Probability Theory and its Applications. Vol. I, Wiley, 1950.

[6] Pinkham, Roger S. An Approach to Linear Inventory-Production Rules. Operations Research 6, March-April 1958.

[7] Vassian, Herbert J. Application of discrete variable servo theory to inventory control. Operations Research 3, May 1955.

[8] Wiener, Norbert. Extrapolation, Interpolation and Smoothing of Stationary Time Series. Wiley 1949.